Worksheet 1 Compression and encryption

Task 1 Compression

Files are compressed using a variety of techniques. An image file is compressed differently than a folder containing text files.

Different compression schemes produce files in a specific format. The extension of a file indicates what this is so that the Operating System opens the file with an application that can decompress it.

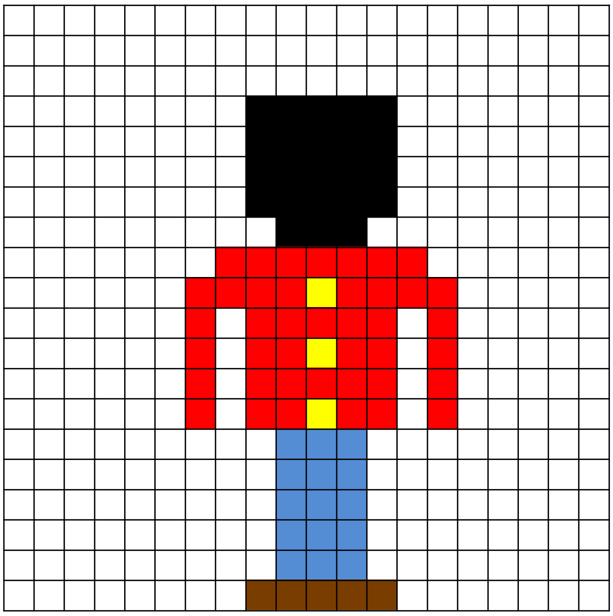
Complete the diagram by matching the file type with the type of compression performed, the description of what it does and the file extension.

Lossless	Lossy
	.JPG
	.MP3
	.MP4
.ZIP	

Description
Only records changes in differences between picture frames of video rather than each entire frame
Removes audio frequencies that the human ear can’t detect
Identifies repeated file content and replaces every occurrence with a reusable code
Compresses pixels in an image by removing colours the human eye cannot distinguish between

Task 2 Calculating compression

A 20x20 bitmap image file is described using a colour depth of 8 bits as shown below:



The file size of this image is calculated by firstly determining the resolution or how many pixels there are: 20 x 20 = 400 pixels.

Each pixel is described by 8 bits so the file size is: 400 pixels x 8 bits = 3,200 bits or 400 bytes.

a) Run length encoding is applied to the file in order to reduce its file size. An encoded sequence of the same data is stored as 2 bytes: the value of the repetition and the colour code. For example the first line would be stored as 20W (the W would be the binary code for white but is summarised here as a character).

For each line show how the data would be encoded using the format: [LengthColour]. The first and fourth lines have been done for you.

(**W**=White, **K**=Black, **R**=Red, **Y**=Yellow, **B**=Blue, **N**=Brown)

Line	Run length encoded sequence
1	20W
2	
3	
4	8W5K7W
5	
6	
7	
8	
9	
10	
11	
12	
13	
14	
15	
16	
17	
18	
19	
20	

b) Determine what the RLE compressed file size is bytes:

c) Express, as a whole number percentage of the original file size, how much compression has been applied to the picture:

d) This type of compression is categorised as lossless. Explain what is meant by this in terms of the picture file:

Task 3 Dictionary compression

A text document contains the proverb:

give a man a fish and you feed him for a day,
teach a man to fish and you feed him for a lifetime

Dictionary compression is applied to the file whereby common **single words and associated space characters** are stored in a separate dictionary component and referenced by an 8 bit binary code.

a) Each character is stored as an 8 bit ASCII code (including spaces). Calculate the storage size of the quote in **bytes**:

b) Complete the dictionary below listing all the repeated words in the order the proverb is presented (note that the underscore represents the space character:

Code	Repeated word
00000000	a_
00000001	man_
00000010	
00000011	
00000100	
00000101	
00000110	
00000111	

c) What will be the size of the dictionary in bytes if each letter of each repeated word is stored in 1 byte?

d) What will be the new size of the quote in bytes if codes from the dictionary are used in place of the repeated words?

e) Express, as a percentage of the original quote size, how much compression has been applied to the text:

f) Dictionary compression can be adjusted to be more effective by storing phrases rather than just single words.

Explain how this concept would significantly reduce the number of bytes used to represent the quote looked at in part (a).

Task 4 Encryption

The Caeser cipher

Messages sent across a network are encrypted using a Caesar cipher.

a) Determine what the encrypted message of “*mary had a little lamb*” would be with a shift of minus 3:

b) Decrypt the short message “fdwfk wkh sljhrq”.

c) The following message has been encrypted using the Caesar cipher but word length and letter cases have been obscured. Use some basic frequency analysis to decrypt the message below: (Punctuation does not shift.)

```
BRXFD QQRWO RVHDK RPLQJ SLJHR Q.LIB RXUKR PLQJS LJHRQ
GRHVQ RWFRP HEDFN ,WKHQ ZKDWB RXKDY HORVW LVDSL JHRQ.
```

The most frequently occurring of letters in the English alphabet are **E**, **T**, **A**, **O**, **I** and **N**.
The least frequent are **Z**, **Q**, **J** and **K**.

Letter	Frequency	Letter	Frequency	Letter	Frequency	Letter	Frequency
A		H		O		V	
B		I		P		W	
C		J		Q		X	
D		K		R		Y	
E		L		S		Z	
F		M		T			
G		N		U			

The Vernam cipher

The Vernam cipher has been proven to be the only cipher that is unbreakable as long as certain rules are followed. These are that the one-time pad must be truly random, used only once and must be hand delivered to the recipient.

a) Use the ASCII code sheet to encrypt the following plaintext: **Rat** with the one-time pad of: **a!H**.

b)

1	0	1	0	0	1	0	**R**
1	1	0	0	0	0	1	**a**

							a
							!

							t
							H

Reverse the XOR operation to decrypt the following ciphertext: **}#<** using a one-time pad of: **5L[**.

c)

1	1	1	1	1	0	1	**}**
0	1	1	0	1	0	1	**5**

							#
							L

							<
							[

Why must the one-time pad be generated from a truly random source rather than being computer generated?

d) Why should a one-time pad only be used once?

e) Why must a one-time pad be hand delivered?